用Excel解決農業研究資料統計分析的方法(一)

敘述統計量

◎農試所農藝組　呂秀英　魏夢麗　呂椿棠

一、前　言

在農業科技研究中，經常要對收集到的資料進行各種統計分析。統計分析是一個極爲繁瑣的過程，若欠缺適當的統計分析軟體，容易因人爲因素的影響，造成分析結果錯誤或誤差較大，準確性無法獲得保證。一些專用統計分析套裝軟體，如SAS(Statistical Analysis System)、SPSS(Statistical Package for the Social Sciences)等，不僅能進行單變數分析，而且可做各種複雜的多變數分析，功能繁多強大、適應性廣，但其操作較爲繁瑣，對一般的生物研究學者而言不易掌握，也不方便使用，再加上它們的價錢通常很昂貴，往往令人怯步。

Excel是美國微軟公司推出的試算表管理軟體，是Microsoft Office 的重要組成部份之一，絕大多數個人電腦用戶都擁有該軟體。Excel以其易學易用的可視化圖形界面，強大的資料處理功能等特點，受到形形色色的用戶青睞。Excel其實不是一個專用統計軟體，但其具有強大的資料管理功能，能製作各種統計圖表，內含豐富的數學、財務會計及統計函數，並且Excel在「工具」中的「資料分析」提供了一組資料分析工具。由於在對農業研究資料進行初步整理時，大多是使用簡單統計描述，繪製各種統計圖表，或者進行基本的統計分析，固然Excel仍有其應用上的限制，用它來進行初步探勘及分析，簡明快捷，易學易用，可使複雜的計算工作變得輕鬆起來。因此，本系列報告在介紹如何使用Excel進行農業科學研究中資料分析的方法及技巧，如敘述統計量、t檢定、F檢定、X^2檢定、變方分析(CRD、RCBD)、相關及迴歸分析。

二、Excel分析資料的途徑及安裝

Excel 內提供3種途徑供資料統計分析用：輸入公式、插入函數、「工具」中的「資料分析」。

假設資料集之位址在A1:A5範圍內，欲求這5筆資料的平均值，則最簡單的方式是使用數學運算符號，如加(＋)、減(－)、乘(×)、除(/)、乘方(^)、括號(())，直接輸入公式，如本例平均值的公式爲=(A1+A2+A3+A4+A5)/5，但這種方式一旦運算公式較複雜或資料量變多時，就顯得非常不實用。第二種分析途徑是使用插入函數的方式，最新版本的Exce1 2003提供了至少80種函數，常用的數學及統計函數有 ：絕對值(ABS)、平方根(SQRT)、個

作　　者：呂研究員秀英
聯絡電話：04-23302301-125

數(COUNT)、頻度分布 (FREQUENCY)、總和(SUM)、算術平均值(AVERAGE)、幾何平均數(GEOMEAN)、最大值(MAX)、最小值(MIN)、中位數(MEDIAN)、眾量(MODE)、平方和(DEVSQ)、標準差(STDEV)、變方(VAR)、峰度(KURT)、歪度(SKEW)等。以本例5筆資料而言，可在資料集以外的某單格內，使用功能表上的「插入」的「函數」選出AVERAGE函數並指定資料範圍，或者直接於單格內輸入=AVERAGE(A1:A5)，可求得其算術平均值。使用者利用這種途徑進行統計分析，必須熟習各種函數名稱，而且一旦需要同時求算多個不同敘述統計量時，就必須給予不同函數名稱，使用上並不方便；而且某些農業研究常見的統計量，如平均值標準差(standard error of mean, SE or SEM)及變異係數(coefficient of variation, CV)，Excel也沒有提供可用的相關函數。

第三種統計分析途徑，則是使用Excel的「工具」中的「資料分析」，此處提供了很多統計分析工具，且克服了上述兩種分析途徑的侷限性。使用這些分析工具時，只需指出資料所在的單格和提供必要的參數，該工具就會使用適宜的統計或工程函數，對資料做處理，給出相應的結果；有些工具在輸出時還能產生圖表，如迴歸分析。

但在安裝Excel時，未必一定加入了「資料分析」，因爲Microsoft Office內設不會自動安裝。若在「工具」中沒有發現「資料分析」，安裝流程如圖一所示。首先要勾選「工具」中的「增益集」，然後按「確定」(圖一，A)；選取了「增益集」後，會有一系列的「增益功能」，從中選擇「分析工具箱」並在旁的方格內加V，然後按「確定」便可完成安裝 (圖一，B)；這時再按「工具」，便會發現「資料分析」一項了(圖一，C)。

三、敘述統計量之求算步驟

進入Excel在「工具」中的「資料分析」，選擇「敘述統計」，可以針對多個變數性狀快速準確地一次計算出各種常用的敘述統計量(descriptive statistics)。例如調查8隻果蠅的翼長(cm)與體重(g)，可在工作表中輸入資料(如圖二，A所示)，然後在「工具」的「資料分析」，選擇「敘述統計」，進入「敘述統計」對話視窗中，操作步驟如下(圖二，A)：

(1) 指定「輸入範圍」－可用輸入範圍位址或用滑鼠拉選範圍；如本例指定的輸入範圍若包含兩變數名稱在內，可直接輸入相對位址A1:B9 (倘使用滑鼠拉選範圍的方式，輸入範圍自動顯示絕對位址$A$1:$B$9)。

(2) 依資料輸入格式選擇「分組方式爲逐欄或逐行」－如本例爲兩性狀分欄輸入，故勾選逐欄。

(3) 勾選「類別軸標記是在第一列上」－因本例在指定範圍時，將兩個性狀的變數名稱(標記)一起納入，故必須勾選此項；但若只有指定資料本身，未將標記納入，則此項毋須勾選。

(4) 在「輸出選項」中勾選分析結果所在位置－可以自行決定要放在同一工作表內某處「輸出範圍」內、「新工作表」或「新活頁簿」；指定「輸出範圍」時，只要隨便輸入一些位址即可，未必要確認輸出結果的內容有多少後才指定符合一定

A　B　C

圖一、Excel「工具」中的「資料分析」之安裝流程：A.進入「工具」的「增益集」，B.勾選「分析工具箱」，C.再進入「工具」就可發現「資料分析」

大小的範圍。如本例希望輸出結果與資料能放在一起，因此勾選輸出範圍並指定範圍位址。

(5) 勾選「摘要統計」－此項爲必要勾選項目。

(6) 平均信賴度、第k個最大值、第k個最小值，爲自由選項，視需要時再勾選－平均信賴度內設爲95%，欲變更時才需勾選此項並修改數值。

(7) 最後按「確定」，即給出輸出結果－如本例結果示如圖二，B 。

四、敘述統計量之輸出結果說明

由於Excel屬於商業用軟體，而且微軟公司是先出版簡體中文版，然後再將其轉換成繁體中文版，因此它所呈現的統計中譯名稱，往往與生物統計領域所慣用的名稱多少有出入，因此有必要加以說明，以免混淆。使用Excel的「工具」中「資料分析」的「敘述統計」，可自動算出平均數(即算術平均值)、標準誤(即平均數的標準差, SEM或簡稱SE)、中間值(即中位數)、眾數、標準差(即SD)、峰度、偏態(即歪度)、範圍、最小值、最大值、總和、個數，如圖二，B所示之本例結果。其中由於果蠅這兩組性狀資料都沒有眾數(即沒有出現頻度最多的值)，Excel以#N/A呈現，表示無法產生眾數。

但Excel無法經由「工具」中「資料分析」自動算出變異係數(CV)。CV的公式爲標準差除以平均數再乘上100，因此可以透過輸入公式求得CV值，以本例而言(圖二，B)，欲求翼長的CV值，只要在某單格內輸入=E7/E3×100 (E7, E3各爲翼長標準差及平均數所在的位址)，同理，可在另一單格內輸入=G7/G3×100 (G7, G3各爲體重標準差及平均數所在的位址)，以求出體重的CV值。

五、CV、SD、SE應用上的區分

CV值的單位爲%，最適合用來比較平均值大小相差很多的資料組，如多個品種的產量變異性差異，以及用來比較測量單位不同的兩組資料的變異性之場合，如

A

B

圖二、用Excel「工具」中的「資料分析」求算果蠅兩組資料的敘述統計量：A.操作畫面及B.輸出結果

不同性狀間之變異性的差異。因此從本例兩性狀的CV值(圖二，B)，可以知道翼長的變異性(CV=44%)大於體重的變異性(CV=28%)。另外，我們也常使用CV值來判斷所調查或測量的樣本是否具有代表性，即資料中是否雜有極端值或不正常值(CV值異常大)。

標準差(SD)和平均值標準差(SE or SEM)的應用場合，最容易被混淆。SD是standard deviation的縮寫，是將變方開平方根後的值 (族群: σ;樣本：s)；而SE是standard error of mean (有些統計教科書上寫成standard deviation of mean)的縮寫，是將樣本標準差(s)除以觀測值數目之開平方根($\sqrt{n}$)後的值，即s/ $\sqrt{n}$，通常在正式學術報告中多以SEM或SE of mean 稱謂 (即多了一個mean，以強調是平均數的標準差，而非觀測值的標準差)，且公式符號以 $S_{\bar{x}}$ 代表，以與樣本標準差SD=s 有所區分。SD和SEM的測量單位都與平均值相同，但SD用來檢視觀測值分布的分散程度，當我們使用mean ±SD時，是要讓讀者可清楚地從這兩個數字中(mean及SD)得到該變數的分布，舉例來說，當資料爲常態分布時，±1 SD表示包含66%的資料，±2 SD表示含有95%的資料，而±3 SD則含有99%的資料。SEM則代表平均值的估計準確度，當我們使用mean ±SEM時，例如A、B兩組資料各爲100 ±2.5及120 ±1.2 (僅限測量單位相同的兩組資料)，表示我們可以推測A組平均值的準確度較B爲低。SEM最適用於測量單位相同的多個平均值間之比較場合，但若是各平均值的測量單位不同，則應改用CV值來比較，CV值較大者表示平均值的變異程度大，即平均值的準確度較低。

六、參考文獻

吳權威、呂琳琳。2004 。Excel 2003函數與統計應用實務。網奕出版社，臺灣臺北。

魏夢麗、陳烈夫、呂秀英。2003 。利用Excel進行農業試驗資料之統計分析的應用及限制。科學農業 51(3,4):63-70 。